

QoS Support for End Users of I/O-intensive Applications using Shared Storage Systems

Xuechen Zhang and Song Jiang
Wayne State University

Kei Davis
Los Alamos National Laboratory

High-performance applications with strict QoS demands include:

- Hurricane forecasting
- Epidemiologic modeling of diseases
- Financial data modeling



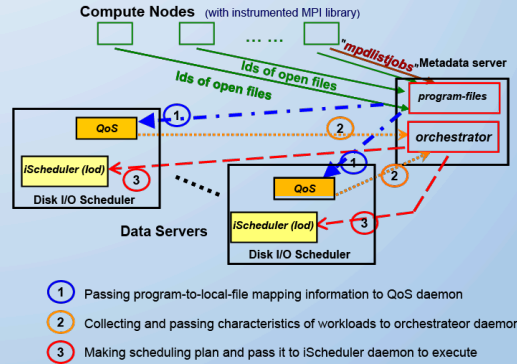
The weaknesses of conventional methods for guaranteeing QoS include:

- Users lack the knowledge or expertise to determine I/O service quality commensurate with apps' required runtimes
- Maintaining a simple performance target may not allow a shared storage system to be efficiently used
- I/O interference among parallel applications can lead to under-utilization of storage devices

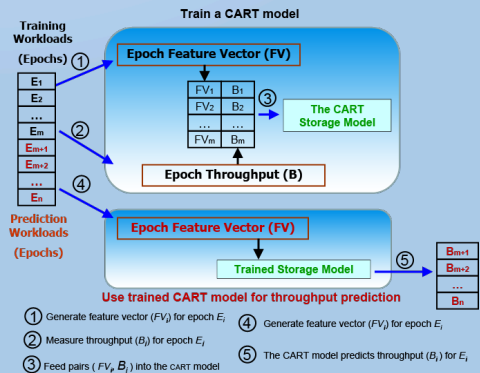
Our Solution: Derived QoS bounds for resource allocations

- Leverage a machine learning technique to automatically convert the end-user QoS requirements into instantaneous derived-throughput bounds at run time
- Schedule requests in each time window to not only meet QoS requirements but also improve data access efficiency on a storage system of multiple data servers
- Efficiently implement the new performance interface on the I/O stack of existing storage systems by instrumenting PVFS2 and MPI-IO

System Architecture

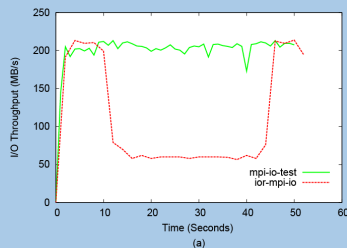


Performance Modeling

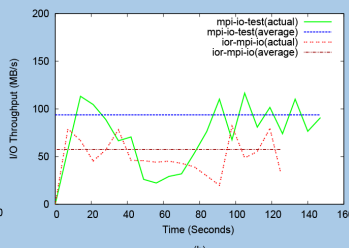


Experimental Set up

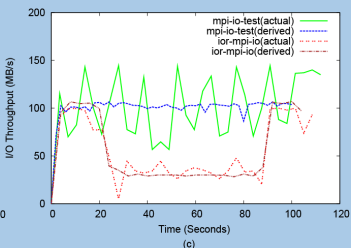
- Nine servers with Dual 1.6-GHz processors, 1 GB memory, 150-GB SATA Disks
- Fedora Linux with CFQ, PVFS2 parallel file system, MPICH2
- Benchmarks: mpi-io-test, ior-mpi-io, ...
- QoS target: 100s for mpi-io-test and 110s for ior-mpi-io in this experiment



Performance W/O Sharing



Throughput Bounds



Derived Bounds (Proposed)